Probabilidades e Aplicações **2020/2021 (1º semestre)**
CC + MAT

UNIVERSIDADE DO MINHO EXAME DE RECURSO (02.02.2021)
Departamento de Matemática Resolução abreviada

1. Numa dada população, certa doença está presente sob forma grave em 3% dos indivíduos, sob forma moderada em 10% e ausente nos restantes 87%. Um exame clínico assinala a presença da doença (i.e., dá resultado *positivo*) em 90% dos casos graves, em 70% dos moderados e em 10% dos saudáveis.

   (a) Um indivíduo escolhido ao acaso na população é submetido ao exame. Sabendo que o resultado foi positivo, qual a probabilidade de que o indivíduo tenha a doença? Explique a sua resolução.

   Para um indivíduo escolhido ao acaso, considerem-se os acontecimentos G [M; A] e +, resp. "o indivíduo tem doença grave [moderada; ausente]" e "o resultado do teste foi positivo". Então, pelo TPT com a partição $\{G, M, A\}$ temos
   $P(+) = P(+|G)P(G) + P(+|M)P(M) + P(+|A)P(A) = .9 \times .03 + .7 \times .1 + .1 \times .87 = 0.184$, . Logo $P(G \cup M \ |+) = 1 - P(A \ |+)$ donde a probabilidade pedida é
   $P(G \cup M \ |+) = 1 - \dfrac{P(+|A)P(A)}{P(+)} = 1 - \dfrac{0.1 \times .87}{.184} = 0.5271739 \simeq 0.527$

   (b) Numa amostra aleatória de 10 indivíduos da população, calcule a probabilidade de "exactamente 7 serem saudáveis e no máximo 1 estar gravemente doente". Explique.

   Sendo X e Y as v.a. que representam o nº de indivíduos da amostra que "não têm a doença" e que "têm a doença sob forma grave", resp., então $(X, Y) \frown M(10; 0.87, 0.03)$ (distribuição trinomial), donde
   $P(X = 7, Y \leq 1) = P(X = 7, Y = 0) + P(X = 7, Y = 1) = 0.08601409 \simeq 0.086$

   ```
   p <- c(0.87, 0.03, 0.10)
   dmultinom(c(7,0,3),prob=p) + dmultinom(c(7,1,2),prob=p)
   ```

2. (i) Se os sismos num determinado local seguem um processo de Poisson à taxa de 6 por ano, qual a probabilidade de haver menos do que 3 sismos em meio ano? E qual a distribuição do intervalo de tempo entre o 1º e o último de 3 sismos consecutivos? Explique a sua resolução.

   Seja N_t o nº de sismos durante t anos; então $N_t \frown Poisson(6t)$, donde para meio ano temos $N_{1/2} \frown Poisson(3)$, logo $P(N_{1/2} < 3) = P(N_{1/2} \leq 2) = 0.4231901 \simeq 0.423$, obtido com o comando `ppois(2, 3)`.
   Como num processo de Poisson (de taxa λ) os intervalos $T_1, T_2, \ldots$, entre acorrências sucessivas são v.a. i.i.d. $Exp(\lambda)$, temos então neste caso $T_1 + T_2 \frown Gama(2, 6)$.

   (ii) Simulou-se no `R` a seguinte amostra de NPA: `rexp(100)*rbinom(100,1,0.6)`. Descreva a v.a. subjacente a esta amostra e refira um fenómeno concreto (do mundo real) que possa ter esta lei de probabilidade.

   A v.a. em causa é o produto XY de duas v.a., $X \frown Exp(1)$ e $Y \frown bi(1, 0.6)$, independentes uma da outra. A v.a. XY vale 0 com probabilidade 0.4 e é positiva – com distribuição $Exp(1)$ – com probablidade 0.6; trata-se portanto de uma mistura de uma v.a. discreta com uma contínua. Exemplo: precipitação atmosférica num dado dia do ano em certo local, supondo que a probabilidade de aí chover nesse dia é 60% e que no caso de chover, a quantidade de precipitação (l/m^2) tem distribuição $Exp(1)$.

3. Alfa e Beto têm cada um uma quantia de 500 € para investir em empresas cotadas na bolsa. Em qualquer destas empresas, o lucro (em euros) ao fim de um ano por cada 10 € investidos é uma v.a. $T = U+V$, sendo U e V v.a. independentes com distribuição uniforme no intervalo $[-\frac{1}{2}, \frac{1}{2}]$. Os lucros em empresas distintas são v.a. mutuamente independentes.

   Alfa investe tudo numa única empresa e Beto investe 10 € em outras 50 empresas (distintas). Represente por $T_0, T_1, \ldots, T_{50}$ os lucros por cada 10 € investidos nas 51 empresas em questão. Seja X o lucro de Alfa e Y o lucro de Beto, ao fim de um ano. Calcule

   (a) o valor médio e a variância de T (justifique);
   $E(U) = E(V) = 0$; $\mathrm{Var}(U) = \mathrm{Var}(V) = \frac{1}{12}$; então $E(T) = E(U+V) = E(U)+E(V) = 0$ e como U e V são independentes, temos $\mathrm{Var}(T) = \mathrm{Var}(U+V) = \mathrm{Var}(U)+\mathrm{Var}(V) = \frac{1}{6}$

   (b) o valor médio e o desvio padrão de X e de Y (justifique);
   Temos $X = 50\,T_0$, donde $\mu_X = E(X) = 50\,E(T_0) = 0$ e $\sigma_X = \sigma_{50\,T} = 50\,\sigma_T = \frac{50}{\sqrt{6}}$.
   Temos $Y = \sum_{i=1}^{50} T_i$, donde $\mu_Y = \sum_{i=1}^{50} E(T_i) = 0$ e como os T_i são v.a. mutuamente independentes, então $\mathrm{Var}(Y) = \sum_{i=1}^{50} \mathrm{Var}(T_i) = 50\,\frac{1}{6}$, donde $\sigma_Y = \sqrt{\frac{50}{6}}$.

   (c) $P(X > 5)$, explicando a sua resolução;
   $P(X > 5) = P(50\,T > 5) = P(T > \frac{1}{10}) = P(U + V > 0.1) = P(V > 0.1 - U) = \frac{1}{2}\,0.9^2 = 0.405$ (é a área do triângulo azul assinalado na figura 1)

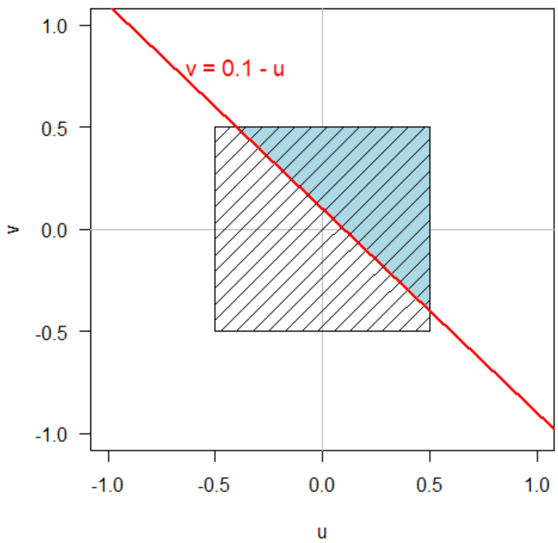

Figura 1: gráfico do quadrado Q (riscas), recta $v = 0.1 - u$ e domínio de integração (triângulo azul)

   (d) um valor aproximado de $P(Y > 5)$ usando (i) o Teorema Limite Central
   Como as v.a. T_i são simétricas, então $n = 50$ é suficiente para que a aproximação pelo TLC seja boa. Logo $P(Y > 5) = P(\sum_{i=1}^{50} T_i > 5) \simeq P(Z^* > 5) = 0.0416$, com $Z^* \frown N\left(0, \sqrt{\frac{50}{6}}\right)$.

```
pnorm(5,0,sqrt(50/6),lower=F)  # 0.04163226
```

   (ii) simulação (explique o raciocínio).Valor aproximado: 0.041 (simulação c/ 10^5 réplicas):

```
r <- 100000; m <- matrix(runif(r*50,-0.5,0.5)+runif(r*50,-0.5,0.5),nr=50);
s <- colSums(m); sum(s>5)/r  # 0.04098
```

   Comente, tendo em atenção a comparação entre as distribuições de X e de Y.

   As distribuições de X e de Y são ambas simétricas em torno de 0, mas $\sigma_X > \sigma_Y$, pelo que $P(X > 5)$ é maior que $P(Y > 5)$; o investimento de Alfa é mais arriscado por ser menos diversificado; a probabilidade de Alfa ganhar [perder] muito é muito maior que a de Beto.

Notas:

(i) A distribuição de X é triangular em $[-50, 50]$ pois corresponde à soma de duas v.a. independentes com distribuição $U[-25, 25]$. As fdp de X e Y (aproximada, i.e., $N\left(0, \sqrt{\frac{50}{6}}\right)$) podem representar-se conforme a figura 2, permitindo concluir graficamente que $P(X > 5)$ é bem maior que $P(Y > 5)$.

(ii) A partir desta fdp triangular pode calcular-se $P(X > 5)$ em alternativa à resolução feita na alínea (c). Trata-se de integrar a fdp de X (a vermelho) no domínio [5, 50], ou seja, calcular a área de um triângulo rectângulo de base 45 e altura $\frac{45}{50^2}$, cujo resultado é $\frac{45^2}{2\times 50^2} = 0.405$

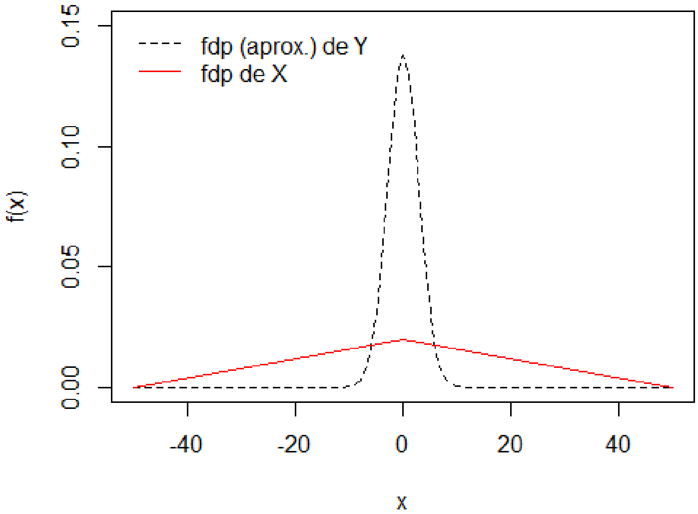

Figura 2: gráfico da fdp de X e da fdp (aproximada pelo TLC) de Y